

Flux networks in metabolic graphs*

Patrick B. Warren, Silvo M. Duarte Queiros, and Janette L. Jones
Unilever R&D Port Sunlight, Bebington, Wirral, CH63 3JW, UK.

A metabolic model can be represented as bipartite graph comprising linked reaction and metabolite nodes. Here it is shown how a network of conserved fluxes can be assigned to the edges of such a graph by combining the reaction fluxes with a conserved metabolite property such as molecular weight. A similar flux network can be constructed by combining the primal and dual solutions to the linear programming problem that typically arises in constraint-based modelling. Such constructions may help with the visualisation of flux distributions in complex metabolic networks. The analysis also explains the strong correlation observed between metabolite shadow prices (the dual linear programming variables) and conserved metabolite properties. The methods were applied to recent metabolic models for *Escherichia coli*, *Saccharomyces cerevisiae*, and *Methanosarcina barkeri*. Detailed results are reported for *E. coli*; similar results were found for the other organisms.

PACS numbers: 87.16.Yc, 87.18.Vf

ABBREVIATIONS

CBM: constraint-based modelling.

CS: complementary slackness (a property of LP solution pairs at optimality).

GAM: growth-associated maintenance (in relation to ATP consumption).

gDW: gram dry weight (referring to biomass).

LP: linear programming.

NGAM: non-growth-associated maintenance (in relation to ATP consumption).

A metabolic network comprises a list of biochemical reactions and their associated metabolites [1]. As such, a convenient representation is in terms of a bipartite graph containing reaction nodes and metabolite nodes, with edges between nodes indicating that a given metabolite is involved in a given reaction [2]. A schematic example is shown in Fig. 1. The metabolic network can be modelled by chemical rate equations, giving the rate of change of the metabolite concentrations in terms of the fluxes, or velocities, of the associated reactions. It is widely accepted though that the metabolism comes to a steady state very quickly, so that the metabolite concentrations are unchanging in time. This means that a flux balance condition holds, and the set of reaction fluxes (the ‘fluxome’) can, essentially, be regarded as the metabolic phenotype. Determination of the fluxome is therefore the focus of considerable theoretical [1], and experimental effort [3, 4]. The global properties of such flux sets have been investigated [5].

When the network is represented as a bipartite graph, the fluxome is traditionally associated with the reaction nodes. Here we show how fluxes can be associated with the *edges* of such a bipartite graph, by combining the reaction fluxes with any metabolite property that is con-

served in the majority of reactions, such as molecular weight. Moreover, assuming the flux balance condition, such an edge-associated flux network is conserved at all the reaction and metabolite nodes apart from a handful of sources and sinks. Thus the edge-associated fluxes resemble, for example, electric currents in a network of resistors [6]. This observation may help with the visualisation of the flow of material in these complex reaction networks. If the flux-balance condition does not hold (for example away from steady state), then the edge-associated flux network can still be constructed provided a set of reaction fluxes is available. In such a case though, the edge fluxes are not in general conserved at the reaction nodes. Finally in the case where the set of reaction fluxes arises from the solution to a linear optimisation or linear programming (LP) problem, such as commonly encountered in constraint-based modelling (CBM), then an edge-associated *yield flux network* can be constructed.

CBM is now a well-established approach for calculating candidate sets of reaction fluxes [1]. It has been applied to micro-organisms from all three domains of life [7, 8, 9, 10], and recently extended to encompass human metabolism [11]. For the growth of micro-organisms a commonly used paradigm has emerged in which the metabolic network is augmented with a biomass reaction consuming the end-points of metabolism in the appropriate ratios, and with exchange reactions to represent the uptake of substrates and the discharge of metabolic by-products. Maximising the flux through the biomass reaction amounts to maximising the specific growth rate of the micro-organism. This approach has been highly successful at predicting the behaviour of micro-organisms [4, 12, 13], and has also been applied to problems in metabolic engineering [14, 15].

As already indicated CBM typically leads to an LP problem for the set of reaction fluxes. Mathematically, every LP problem has an associated dual [16, 17]. The dual variables are known as shadow prices reflecting an economic interpretation of the dual problem. In CBM, shadow prices were first investigated by Varma and Palsen who showed that they are, essentially, yield coeffi-

*Published in Phys. Biol. **6**, 046006 (2009).

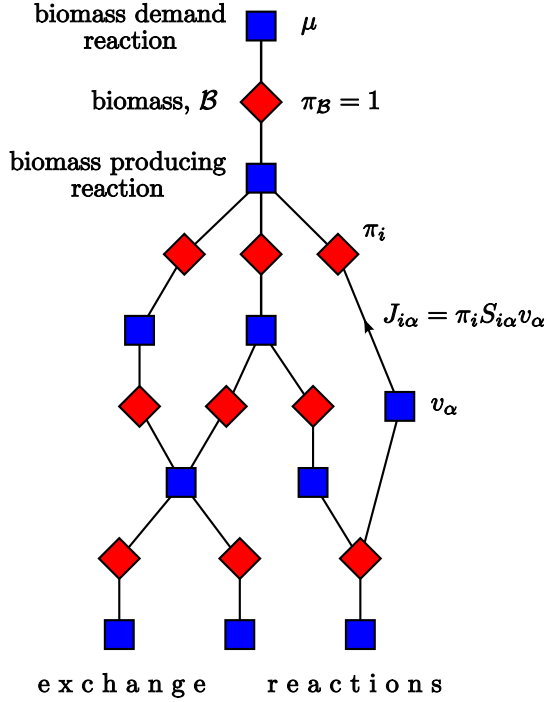


FIG. 1: Schematic metabolic network as a bipartite graph. Edges go from reaction nodes (\square , blue) to metabolite nodes (\diamond , red). Reaction fluxes, v_α , metabolite shadow prices, π_i , and the stoichiometry matrix, $S_{i\alpha}$, combine to make the edge-associated yield flux network, $J_{i\alpha}$. It can be shown (see text) that the $J_{i\alpha}$ are conserved at all nodes except a limited number of exchange reactions which serve as sources, and the biomass demand reaction which serves as a sink. The biomass shadow price is unity, and the flux through the biomass demand reaction is the specific growth rate μ .

cients [18, 19]. Later, the dual problem was explicitly formulated by Burgard, Maranas, and coworkers, for use in multi-level optimisation problems [15, 20]. Recently we described a thermodynamic interpretation of the dual problem [21]. The aforementioned yield flux network can be generated very naturally by combining a candidate set of reaction fluxes with the corresponding set of shadow prices. The properties of the yield flux network (such as flux conservation) follow from the so-called complementary slackness (CS) relations. The parallel construction of the various flux networks explains the strong correlation between shadow prices and conserved metabolite quantities such as molecular weight.

I. METHODS

A. Conserved edge-associated flux networks

Mathematically, a metabolic network is conveniently described by the stoichiometry matrix $S_{i\alpha}$, giving the number of moles of the i -th metabolite consumed or pro-

duced by the α -th reaction. We make a distinction between balanced ‘internal’ reactions which typically represent biochemical transformations or membrane transport processes, and imbalanced reactions introduced in CBM such as the biomass reaction or the exchange reactions. We suppose there are $\alpha = 1 \dots R$ reactions (of all types) and $i = 1 \dots M$ metabolites, with typically $M < R$. The convention we adopt is that $S_{i\alpha}$ is positive for products, negative for reactants. The corresponding bipartite graph has R reaction nodes and M metabolite nodes (Fig. 1). An edge connects a reaction node to a metabolite node if and only if $S_{i\alpha} \neq 0$. We additionally suppose the bipartite graph is directed and adopt the convention that all edges start at reaction nodes and end on metabolite nodes.

In terms of the stoichiometry matrix, the chemical rate equations are

$$\frac{dc_i}{dt} = \sum_{\alpha} S_{i\alpha} v_{\alpha} \quad (1)$$

where c_i is the concentration of the i -th metabolite, and v_{α} is the flux through the α -th reaction (reaction velocity). The reaction fluxes are typically measured in units of mol/gDW.hr where gDW means gram dry weight of biomass. In steady state, $dc_i/dt = 0$, leading to the flux balance condition

$$\sum_{\alpha} S_{i\alpha} v_{\alpha} = 0. \quad (2)$$

Let q_i be any property of the i -th metabolite which is conserved in the internal reactions, for example molecular weight, number of atoms, and so on. Then the corresponding edge-associated flux network is defined by

$$Q_{i\alpha} = q_i S_{i\alpha} v_{\alpha} \quad (3)$$

(no implied summation). For example if q_i is molecular weight we generate what we term the mass flux network, if q_i is the number of atoms of a given element we generate an elemental flux network, and so on. From our sign convention for the stoichiometry matrix one has $Q_{i\alpha} > 0$ for edges connected to product metabolite nodes, assuming $q_i > 0$ and $v_{\alpha} > 0$. Hence, typically, the flux flows from reactants to products.

The flux network $Q_{i\alpha}$ defined in Eq. 3 is conserved at all the metabolite nodes since

$$\sum_{\alpha} Q_{i\alpha} = q_i (\sum_{\alpha} S_{i\alpha} v_{\alpha}) = 0, \quad (4)$$

using the flux balance condition Eq. 2. Moreover $Q_{i\alpha}$ is conserved at all the internal reaction nodes since

$$\sum_i Q_{i\alpha} = (\sum_i q_i S_{i\alpha}) v_{\alpha} = 0 \quad (5)$$

where, by definition, $\sum_i q_i S_{i\alpha} = 0$ expresses the fact that q_i is conserved in the α -th reaction. However, $Q_{i\alpha}$ may not necessarily be conserved at imbalanced reaction nodes since $\sum_i q_i S_{i\alpha}$ is not necessarily zero. Clearly these are important nonetheless, since they represent sources

and sinks for the $Q_{i\alpha}$. As discussed in more detail below, in a typical metabolic model these imbalanced reaction nodes are restricted to a small set of so-called exchange and demand reactions (see Fig. 1).

In some sense these edge-associated flux networks, and the yield flux network introduced below, are more fundamental than the set of reaction fluxes. By construction they are insensitive to either local rescaling of the reaction stoichiometry ($S_{i\alpha} \rightarrow S_{i\alpha} \times r_\alpha$ and $v_\alpha \rightarrow v_\alpha/r_\alpha$), or local coarse-graining of metabolites ($S_{i\alpha} \rightarrow S_{i\alpha} \times r_i$ and $q_i \rightarrow q_i/r_i$). Since these rescalings apply locally, like a gauge transformation, one might say the edge-associated flux networks are gauge invariant.

B. The linear programming problem in constraint-based modelling

As outlined in the introduction, the typical application of CBM leads to a so-called primal LP problem, with a corresponding dual LP problem. The details are discussed in the following subsections.

Primal linear programming problem : The variables in the primal LP problem are the reaction fluxes v_α . The constraints are the flux balance conditions in Eq. 2, augmented by additional constraints as follows. Firstly, thermodynamic considerations may lead to some of the internal reactions being judged to be irreversible in which case $v_\alpha \geq 0$. In principle the reaction fluxes can also be ‘capped’ ($v_\alpha^{\min} \leq v_\alpha \leq v_\alpha^{\max}$) or fixed to a prescribed value ($v_\alpha = v_\alpha^{\text{fix}}$) although in practice this is rarely done.

Next, exchange / demand reactions represent the uptake or discharge of substrates from the environment. By convention they are imbalanced reactions of the form $M_i \rightleftharpoons \emptyset$, where M_i is a metabolite, so a positive flux represents discharge of the corresponding substrate and negative flux represents uptake from the environment. For these reactions $S_{i\alpha} = -1$. There is no distinction between demand and exchange reactions, although we tend to restrict the phrase ‘demand reaction’ to the case when the flux is expected to be positive. Exchange / demand reactions are classified according to the allowed flux range, as follows :

$$v_\alpha \begin{cases} = 0 & \text{(closed),} \\ \geq 0 & \text{(half-closed; uptake prevented),} \\ \geq -v_i^{\max} & \text{(open to a limited extent),} \\ \text{unconstrained} & \text{(fully-open).} \end{cases} \quad (6)$$

In the third of these, v_i^{\max} is the maximum specific uptake rate of the given substrate. Most of exchange reactions are half-closed, since there is a need to prevent arbitrary uptake. For certain essential minerals, dissolved gases, nutrients, and vitamins, the exchange reactions are fully open so that the corresponding substrates can be freely taken up or discharged by the organism. In typical applications, one or two exchange reactions are also opened to a limited extent, representing growth-limiting substrates such as carbon/energy sources.

The biomass reaction consumes the end-points of metabolism such as amino acids, nucleotides, lipids, and co-factors. In an extension to the usual paradigm, we split this into an irreversible biomass producing reaction $\sum_i b_i M_i \rightarrow \mathcal{B}$ and an open biomass demand reaction $\mathcal{B} \rightleftharpoons \emptyset$, where \mathcal{B} is a artificial metabolite representing biomass. The stoichiometry coefficients b_i typically have units mol/gDW. The flux through the biomass demand reaction is the specific growth rate, μ , typically with units 1/hr. Since \mathcal{B} only features in these two reactions, the irreversibility of the production step ensures $\mu \geq 0$.

Finally, the energetic requirements of the organism are taken care of by including growth associated maintenance (GAM) and (for high accuracy work) non growth associated maintenance (NGAM) reactions. Again in an extension to the common paradigm, we account for these by including an irreversible energy-generating reaction $\text{ATP}^{4-} + \text{H}_2\text{O} \rightarrow \text{ADP}^{3-} + \text{H}^+ + \text{HPO}_4^{2-} + \mathcal{E}$ where \mathcal{E} is a artificial metabolite representing the energy that can be gained by hydrolysing (one mole of) ATP. For the GAM, \mathcal{E} is included amongst the metabolites consumed in the biomass producing reaction with the coefficient $b_\mathcal{E}$ representing the GAM requirement. For the NGAM, we add an energy demand reaction $\mathcal{E} \rightleftharpoons \emptyset$ with a positive lower bound for the flux, $v_\mathcal{E} \geq v_\mathcal{E}^{\min}$ where $v_\mathcal{E}^{\min}$ represents the NGAM requirement.

This completes the specification of the primal LP problem in typical applications of CBM. The constraints on v_α specify a so-called feasible solution space. The aim is to maximise the specific growth rate, μ , whilst remaining within the feasible solution space. The introduction of \mathcal{B} and \mathcal{E} allows us to move the non-trivial flux bounds and the target of the optimisation to demand reactions, with a corresponding simplification to the dual problem. Numerically, solutions to this LP problem can be found by a straightforward application of LP techniques, for example the simplex algorithm [16, 17]. A MATLAB toolbox for solving LP problems in CBM has been released by the Palsson group [26]. For the present study, we used a bespoke interface to the GNU LP kit (GLPK), which provides an efficient implementation of the simplex algorithm.

Dual linear programming problem : Now we turn to the dual LP problem. The dual variables are shadow prices associated with constraints in the primal LP problem. In particular the flux-balance conditions in Eq. 2 generate a set of shadow prices π_i for the metabolites. The shadow prices in the dual problem are then subject to constraints that correspond to the variables (reaction fluxes) in the primal problem. As mentioned above, the shadow prices can be interpreted as yield coefficients [18, 19]. A full derivation of the construction rules presented below is given in the Appendix.

For the internal reactions, the constraints on the π_i can be written in terms of the derived quantities [21]

$$B_\alpha = \sum_i \pi_i S_{i\alpha}. \quad (7)$$

These are defined for all reactions (the resemblance to

reaction affinity will be discussed shortly). In the dual problem, each reversible or irreversible internal reaction generates a constraint on the B_α according to :

$$B_\alpha \begin{cases} = 0 & (\text{reversible, } v_\alpha \text{ unconstrained}), \\ \leq 0 & (\text{irreversible, } v_\alpha \geq 0). \end{cases} \quad (8)$$

This result applies only in the two cases indicated. The generalisation to more complicated situations, such as fixed fluxes and double-bounded fluxes, is given in the Appendix.

The constraints arising from the internal reactions are supplemented by additional constraints arising from the exchange and demand reactions. Since these reactions involve only one metabolite ($M_i \rightleftharpoons \emptyset$ with $S_{i\alpha} = -1$), the corresponding constraint simplifies to feature only the corresponding metabolite shadow price. The constraints associated with the exchange / demand reactions are :

$$\pi_i \begin{cases} = 0 & (\text{fully open, } v_\alpha \text{ unconstrained}), \\ \geq 0 & (\text{half-closed, } v_\alpha \geq 0, \\ & \text{or limited, } v_\alpha \geq -v_i^{\max}), \\ \text{unconstrained} & (\text{closed, } v_\alpha = 0). \end{cases} \quad (9)$$

The biomass demand reaction is special and the corresponding constraint is

$$\pi_B = 1 \quad (\text{biomass}). \quad (10)$$

This arises because it is the flux through the biomass demand reaction that is the optimisation target in the primal LP problem. It also make sense since by definition the biomass yield coefficient for adding more biomass is unity. Given that $\pi_B = 1$ is dimensionless, the units of π_i for the other metabolites are typically gDW/mol (inverse to the units of b_i). Again this makes sense in terms of the π_i being yield coefficients.

The objective function in the dual LP problem is to minimise $w = \sum_i \pi_i v_i^{\max}$, where the sum is over the limited exchange reactions only. One can show from the strong duality theorem [16] that the minimum value of w is equal to the maximum value of the specific growth rate μ , provided both problems have solutions. It is quite common that there is only one limited exchange reaction, representing single-substrate limitation. In this case the dual objective can be taken to minimise the shadow price of the corresponding metabolite. It follows that v_i^{\max} does not enter the dual problem any more, and the metabolite shadow prices are independent of growth rate. Also, at optimality one has $\mu = w = \pi_i v_i^{\max}$. But μ/v_i^{\max} is the standard definition of the yield coefficient, confirming the interpretation of π_i as the yield coefficient for the limiting substrate.

This construction of the dual LP problem extends and simplifies the results presented in [21]. It also essentially recovers the results obtained by Burgard and Maranas and coworkers [15, 20]. Numerically, the dual problem can of course be solved directly, however the shadow prices are often obtained ‘for free’ as a by-product of

the primal LP solution method. This is the case with the simplex algorithm for instance [16].

Complementary slackness relations : At optimality a number of so-called complementary slackness (CS) relations hold, linking the solutions to the primal and dual LP problems. These are also derived in the Appendix. For irreversible internal reactions the CS relation is $B_\alpha v_\alpha = 0$. There is no CS relation for reversible internal reactions but, since $B_\alpha = 0$ is imposed, it follows that at optimality $B_\alpha v_\alpha = 0$ for all internal reactions. For the exchange reactions the CS relations are $\pi_i v_\alpha = 0$ for half-closed exchange reactions and $\pi_i (v_\alpha + v_i^{\max}) = 0$ for the limited exchange reactions. There is no CS relation for fully open exchange reactions but again, since $\pi_i = 0$ is imposed (apart from biomass), it follows that at optimality $\pi_i v_\alpha = 0$ holds for all exchange reactions except for limited exchange reactions operating at the lower flux bound ($v_\alpha = -v_i^{\max}$). There is no CS relation for the biomass demand reaction since it is assumed fully open.

C. The yield flux network

We now show how a pair of complementary solutions to the above primal and dual LP problems can be used to construct a naturally conserved edge-associated flux network, similar to those constructed in section I A. We start by defining the quantities

$$J_{i\alpha} = \pi_i S_{i\alpha} v_\alpha \quad (11)$$

which typically have units of 1/hr. The $J_{i\alpha}$ are conserved at metabolite nodes since

$$\sum_\alpha J_{i\alpha} = \pi_i (\sum_\alpha S_{i\alpha} v_\alpha) = 0 \quad (12)$$

follows from the flux balance condition. At optimality the $J_{i\alpha}$ are also conserved at internal reaction nodes since

$$\sum_i J_{i\alpha} = (\sum_i \pi_i S_{i\alpha}) v_\alpha = B_\alpha v_\alpha = 0, \quad (13)$$

from complementary slackness.

Exchange reaction nodes are linked by a single edge to the corresponding metabolite. For these edges, $S_{i\alpha} = -1$ and $J_{i\alpha} = -\pi_i v_\alpha$. Complementary slackness therefore implies $J_{i\alpha} = 0$ unless the exchange reaction happens to be operating at the lower flux bound in which case $J_{i\alpha} = \pi_i v_i^{\max}$. For the biomass demand reaction one has $J_{i\alpha} = -\mu$ since $v_\alpha = \mu$ and $\pi_B = 1$.

Thus we conclude that at optimality the yield fluxes $J_{i\alpha}$ are conserved at all nodes of the bipartite graph, except for exchange reaction nodes where the reaction happens to be operating at the lower flux bound, which act as sources, and the biomass demand reaction node, which acts as a sink. We argue this justifies the notion that the $J_{i\alpha}$ constitute a ‘yield flux network’ indicating how material which contributes to growth is transmitted through the network. If there is only one limited

exchange reaction, the conservation law derived above implies $\pi_i v_i^{\max} = \mu$. But at optimality the strong duality theorem shows this is true, as we have discussed in section IB.

D. Comparison between flux networks

The yield flux network $J_{i\alpha}$ and the flux networks $Q_{i\alpha}$ introduced in section IA are very similar. They share the *same* small subset of reaction nodes which act as sources and sinks, namely the exchange and demand reaction nodes (Fig. 1). Thus one might expect the pattern of fluxes to be similar. Moreover, since $J_{i\alpha}$ and $Q_{i\alpha}$ are constructed from identical stoichiometry coefficients and reaction flux sets (see Eq. 3 and Eq. 11), this suggests that one would expect a strong correlation between the shadow prices π_i and conserved metabolite properties q_i . This is confirmed by studies of genome-scale metabolic reconstructions, discussed in section II. We cannot provide a proof of an exact relationship between the π_i and q_i , and indeed the correlation is not expected to be perfectly linear since the corresponding source terms are not necessarily in exact proportion. For example the yield flux network will typically have a single exchange reaction node as a source node (*i. e.* the one operating at the lower flux bound), whereas the mass flux network has sources at all the exchange reaction nodes which carry a reaction flux (since all metabolites have some molecular weight). By the same argument, it is of course not surprising that different conserved molecular quantities are only approximately linearly correlated, for example, molecular weight is only approximately proportional to atom count.

E. Analogies to chemical thermodynamics

Recently we described a thermodynamic interpretation of the dual problem [21]. For completeness, we summarise the analogies and differences between the dual LP problem and non-equilibrium thermodynamics as applied to these networks by Beard and coworkers [22, 23, 24]. There is obviously an analogy between Eq. 7 and Eq. 8, and conventional chemical thermodynamics [25], wherein

$$\begin{aligned} \pi_i &\leftrightarrow \text{chemical potential,} \\ B_\alpha &\leftrightarrow \text{reaction affinity.} \end{aligned} \quad (14)$$

However the analogy fails to be exact since the CS conditions require, at optimality, $B_\alpha v_\alpha = 0$. This means that whenever there is a flux through a reaction ($v_\alpha \neq 0$) the corresponding ‘affinity’ vanishes ($B_\alpha = 0$). This stands in sharp contrast to conventional non-equilibrium thermodynamics where a flux through a reaction is usually associated with a negative reaction affinity (driving force) [22, 23, 24]. It shows that the thermodynamic interpretation of the dual problem cannot be put into exact

correspondence with conventional non-equilibrium thermodynamics.

F. Genome-scale metabolic reconstructions

A genome-scale metabolic reconstruction encompasses all the biochemical transformations allowed for by enzymes encoded on the genome of the organism of interest. As such it represents the entire metabolic capability of the organism. A growing number of such reconstructions are becoming available. The principal genome-scale model used in the present study is iAF1260 for *Escherichia coli* [10], which is perhaps the most complete whole-organism model currently available. We have also studied iND750 for *Saccharomyces cerevisiae* [8], and iAF692 for the archeal methanogen *Methanosarcina barkeri* [9], in addition to an earlier model iJR904 for *E. coli* adjusted slightly to account for later literature [7, 27, 28].

The energetic requirements in these models are represented by a GAM component in the biomass producing reaction and a separate NGAM reaction. For our calculations we retain the GAM requirement, but the NGAM requirement was turned off for simplicity ($v_E^{\min} = 0$). We have checked that this approximation has little influence on our results. Exchange reactions are provided for all the extracellular metabolites in these models. By default they are half-closed, meaning the flux is constrained to be non-negative so discharge only is possible. A subset of the exchange reactions are made fully open, on a case-by-case basis (details available on request). In our calculations *one* additional exchange reaction was also opened to a limited extent representing the limited availability of a substrate under single-substrate growth limitation conditions.

For genome-scale problems it is often the case that even at optimality the flux distribution is still not uniquely constrained, because there may be alternate pathways in the metabolism. Mathematically this shows up in the existence of alternate optima in the LP problem [29]. This is an interesting phenomenon which somewhat complicates the analysis. If the primal LP problem has alternate optima, there will be a similar plurality of solutions to the dual problem. However for every optimum solution to the primal problem, there exists a complementary optimal solution to the dual problem. For instance the shadow prices generated ‘for free’ by the simplex algorithm are automatically complementary to the primal solution. It is important to note that the yield flux network described above must be constructed using complementary solution pairs, since the CS relations apply only in this case. The genome-scale models discussed below exhibit the phenomenon of alternate optima. However we have checked rather carefully that we have used representative solution pairs in presenting our results.

G. Statistical analysis

We undertook statistical analysis of the shadow price distributions for selected conditions and organisms although the results are somewhat inconclusive. We attempted to fit the observed distributions, using maximum likelihood estimators, to a log-Normal, a χ distribution, an inverted χ distribution, and a general distribution $\sim \exp[-(\pi_i^2 + \omega\beta)/\omega\pi_i]$ which has tunable exponential asymptotic behaviour for large and small π_i (ω and β are parameters). However none of these distributions could be said to fit the observed distributions, as judged by the Kolmogorov-Smirnov test [30]. Aside from attempting to fit the full distribution, we also determined whether the observed distributions were compatible with power-law asymptotic behaviour at large and small π_i , using tail estimators of Hill [31] and Meerschaert-Scheffer [32]. Our results gave exponents systematically greater than 3 (see for example Fig. 2, upper plot), which is generally taken to be an indication of exponential asymptotic behaviour rather than power-law behaviour. When we apply these tail estimators to the distribution for $|J_{i\alpha}|$, we recover an exponent value $\approx -3/2$ for large magnitudes, shown as the dashed line in figure 2(b).

II. RESULTS AND DISCUSSION

We report results for iAF1260 for *E. coli* [10], under various conditions. Similar results were obtained for the other organisms and models studied. The data used to generate figures 2–4 has been compiled into an Excel spreadsheet, and is given as supplementary material.

We first discuss the statistical distribution of the shadow prices and yield fluxes. Figure 2(a) shows that the shadow prices in iAF1260 have a broad distribution of around three orders of magnitude. We used statistical tests described in section I G to analyse the distribution, however these were rather inconclusive. We have concluded though that there is unlikely to be any asymptotic power-law behaviour in the distributions. For the cases studied, $\approx 80\%$ or more of the metabolites have a positive shadow price, $\approx 15\%$ have a zero shadow price meaning that the growth rate is unchanged if the provision of these metabolites is altered, and $\approx 5\%$ or less have a negative shadow price meaning the growth rate actually goes down if that metabolite is injected into the system. Figure 2(b) shows the distribution of the yield fluxes $J_{i\alpha} = \pi_i S_{i\alpha} v_\alpha$. This distribution *does* appear to show asymptotic power-law behaviour for large magnitudes. It is notable that the exponent appears to be the same as has been found for the reaction flux distribution [5]. This is interesting since the yield flux network is gauge invariant in the sense discussed at the end of section I A, whereas the reaction fluxes are not gauge invariant. The fact that we observe power-law behaviour in the yield flux network therefore strengthens the earlier analysis of [5].

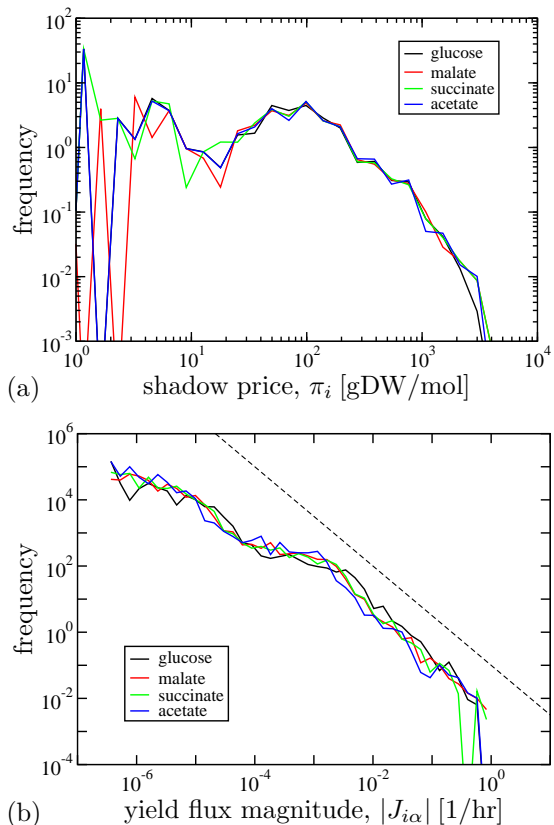


FIG. 2: Distribution of (a) shadow prices and (b) yield fluxes for *E. coli* (iAF1260) growing on various substrates under aerobic conditions. The dashed line in (b) is the power-law $|J_{i\alpha}|^{-3/2}$.

Now we turn to the correlation between shadow prices and conserved molecular properties. Figure 3 shows the shadow price as a function of metabolite formation free energy, molecular weight, and total atom count. To obtain these plots, metabolite formation free energies (where available) and molecular weights are taken from [10], and the atom count is computed from the atomic formulae in [10]. The weakest correlation is with (minus) the free energy of formation. This is unsurprising since the formation free energy is imperfectly conserved in reactions. A stronger correlation is found with molecular weight and the strongest correlation is with atom count. These quantities are conserved since reactions in these genome-scale models are charge- and mass-balanced. The shadow price is more strongly correlated with atom count than molecular weight because there is less spread in the magnitude of the values for atom count. The shadow prices discussed here are ‘molar’ yield coefficients. One can of course define a ‘mass’ yield coefficient by dividing by the molecular weight. The dashed line in figure 3(b) shows that the mass yield coefficients are approximately constant with a value of ≈ 0.5 gDW/g (we have drawn back from undertaking a linear regression analysis as we believe this would over-interpret the data

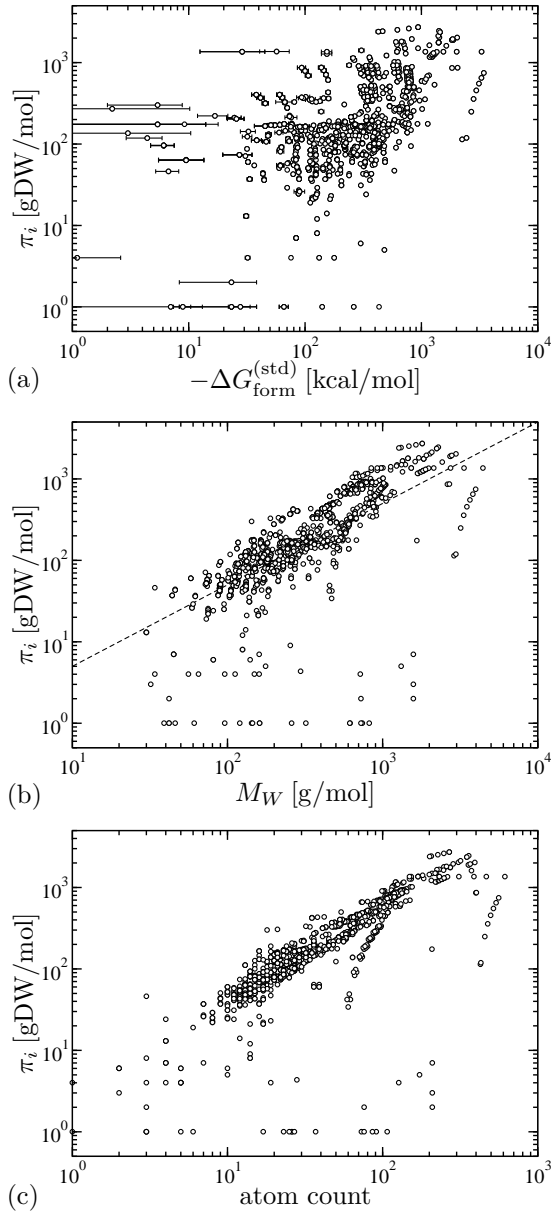


FIG. 3: Shadow prices for *E. coli* (iAF1260) growing on glucose under aerobic conditions as a function of (a) (minus) metabolite formation free energy, (b) molecular weight, and (c) total atom count. The dashed line in (b) is $\pi_i = 0.5 \times M_W$ corresponding to a mass yield coefficient of 0.5 gDW/g.

and would not add any new insights).

The use of shadow prices to measure efficiencies in a model of the central metabolism of *E. coli* was pioneered by Varma and Palsson [18, 19]. Our calculations extend the scope of this analysis to more recent genome-scale metabolic models. Figure 2 and figure 4(a) shows the shadow prices for *E. coli* grown on four different limiting carbon/energy sources. Despite the spread in growth rates from these sources, by and large there is little difference in the shadow price distribution. Invoking the

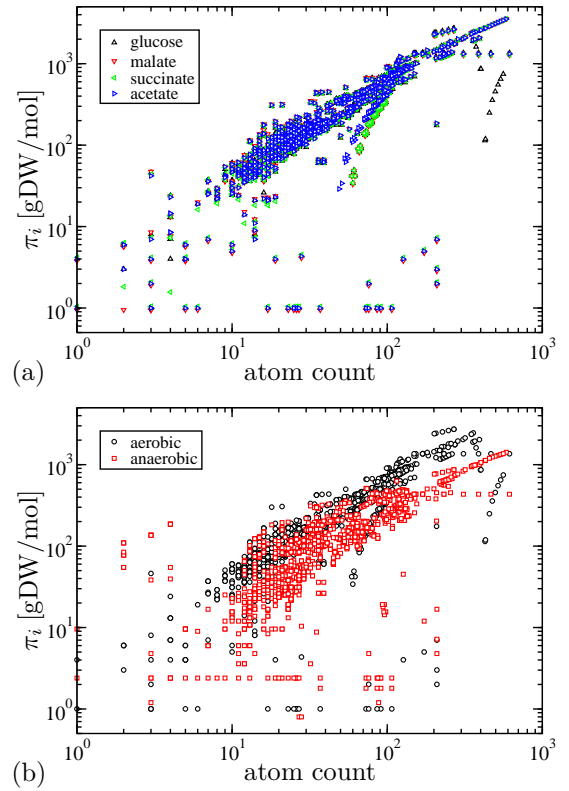


FIG. 4: Shadow prices for *E. coli* (iAF1260) growing (a) on various substrates under aerobic conditions, and (b) on glucose under aerobic and anaerobic conditions.

efficiency arguments of Varma and Palsson, this plot suggests that the metabolic network of *E. coli* has evolved to be equally efficient for growing on a variety of substrates. This reflects the ‘bow-tie’ structure of the metabolic network [33, 34, 35], as substrates are first broken down to a dozen or so common precursors, before being re-assembled into the components required for growth.

Figure 4(b) shows a significant overall lowering of the shadow prices for anaerobic growth compared to aerobic growth. This reflects a reduced efficiency of the network, as more effort has to go into satisfying the energetic requirements of the organism in the absence of oxidative phosphorylation. Two further calculations support this conclusion (data not shown). Firstly, a similar reduction in shadow prices is found for growth under aerobic conditions with the ATP synthase reaction disabled. Secondly, if the NGAM energy demand reaction is thrown fully open ($\mathcal{E} \rightleftharpoons \emptyset$) so that the organism can trivially satisfy its energy requirements, the shadow prices are practically unchanged on going from aerobic to anaerobic conditions.

III. CONCLUSION AND OUTLOOK

To summarise, we have examined the problem of constructing conserved flux networks defined on the edges of

bipartite metabolic graphs. We find that such networks can be generated by combining a conserved metabolic property such as molecular weight, with the reaction fluxes. A similarly conserved, edge-associated flux network can be constructed from a natural combination of the primal and dual solutions to the LP problem that typically arises in CBM. The correspondence between these edge-associated flux networks is responsible for the high correlation between shadow prices and conserved molecular properties. The construction of these networks opens the way for further investigations, both of the global properties of the flux distribution (Fig. 2) [5], and the scaling theory of transport in complex networks [6].

—oOo—

SMDQ acknowledges support by the European Commission through the Marie Curie Transfer of Knowledge project BRIDGET (mktd-cd 2005029961)

APPENDIX

This appendix presents for completeness a derivation of the dual LP problem and the CS relations for the typical LP problem that arises in CBM. It is based on the development in §6.5 in [16] (see also [17]). In this approach we assign a Lagrange multiplier to each constraint in the primal problem, for example the shadow prices are multipliers associated with the flux balance constraints of Eq. 2. Flux constraints are handled by conversion to quadratic equalities. Thus, restricting the analysis for the time being to the two most common flux constraints, we have

$$\begin{aligned} v_\alpha \geq 0 &\Rightarrow v_\alpha = u_\alpha^2, \\ v_\alpha \geq -v_i^{\max} &\Rightarrow v_\alpha + v_i^{\max} = u_\alpha^2. \end{aligned} \quad (\text{A15})$$

The first corresponds to irreversible internal reactions and half-closed exchange reactions. The second corresponds to limited exchange reactions. Adopting the Lagrange multiplier approach, we replace the original constrained linear optimisation problem by the following problem in which we seek the unconstrained maximum of

$$\begin{aligned} Z = \mu + \sum_{i\alpha} \pi_i S_{i\alpha} v_\alpha + \sum'_\alpha y_\alpha (v_\alpha - u_\alpha^2) \\ + \sum''_\alpha y_\alpha (v_\alpha + v_i^{\max} - u_\alpha^2) \end{aligned} \quad (\text{A16})$$

where the first term is the original objective function μ , the second term incorporates the flux balance conditions with π_i being Lagrange multipliers, and the third and fourth terms accommodate the quadratic equalities with y_α being the multipliers. The prime and double prime restrict the sums to the respective reactions with a zero or non-zero lower flux bound. We rewrite this as

$$\begin{aligned} Z = \sum''_\alpha y_\alpha v_i^{\max} + \mu + \sum_\alpha B_\alpha v_\alpha \\ + \sum'_\alpha y_\alpha v_\alpha + \sum''_\alpha y_\alpha v_\alpha \\ - \sum'_\alpha y_\alpha u_\alpha^2 - \sum''_\alpha y_\alpha u_\alpha^2 \end{aligned} \quad (\text{A17})$$

where $B_\alpha = \sum_i \pi_i S_{i\alpha}$ is introduced as a definition to correspond to the main text.

From Eq. A17 one condition that Z is an extremum is $\partial Z / \partial v_\alpha = 0$, implying

$$B_\alpha = \begin{cases} 0 & (\text{unbound}), \\ -y_\alpha & (\text{bound}), \\ -1 & (\text{biomass}) \end{cases} \quad (\text{A18})$$

(‘unbound’ for unbounded reactions; ‘bound’ for reactions with a zero or non-zero lower flux bound, ‘biomass’ for the biomass demand reaction which we assume is unbounded). We notice from Eq. A17 that $y_\alpha \geq 0$ is required for Z to be a *maximum*, otherwise $Z \rightarrow \infty$ as $u_\alpha \rightarrow \pm\infty$. From this and the second of Eq. A18 we deduce that $B_\alpha \leq 0$ for reactions with a lower flux bound. Taken with the first of Eq. A18 this gives the B_α -conditions for the internal reactions quoted in the main text. For the exchange reactions, one has $S_{i\alpha} = -1$ for the metabolite involved in the reaction, and $S_{i\alpha} = 0$ for every other metabolite. Hence $B_\alpha = -\pi_i$. From this, and Eq. A18 and $y_\alpha \geq 0$, we recover the conditions on the exchange reaction metabolite shadow prices quoted in the main text.

The CS relations follow from $\partial Z / \partial u_\alpha = 0$. This implies $y_\alpha u_\alpha = 0$, and hence $B_\alpha v_\alpha = 0$, or $B_\alpha (v_\alpha - v_\alpha^{\min}) = 0$, for reactions with a zero, or non-zero, lower flux bound respectively. Expanding this to the various cases gives the CS relations quoted in the text.

Still following the development in [16], we can read off from Eq. A17 that the dual objective function is to minimise $w = \sum''_\alpha y_\alpha v_i^{\max}$. But $y_\alpha = -B_\alpha = \pi_i$ for exchange reactions, hence $w = \sum'_i \pi_i v_i^{\max}$, as quoted in the main text.

This whole approach can readily be extended to other classes of reactions. For example a reaction with a prescribed flux leads to the corresponding B_α being unrestricted and a contribution $B_\alpha v_\alpha^{\text{fix}}$ being added to w where v_α^{fix} is the fixed flux value. As a special case of this, a reaction which is disabled ($v_\alpha^{\text{fix}} = 0$) has the corresponding B_α made unrestricted in the dual problem. As another special case, an exchange reaction with a specified flux has a contribution $-\pi_i v_\alpha^{\text{fix}}$ added to w . It follows that $\pi_i = -\partial \mu / \partial v_\alpha^{\text{fix}}$. This proves the shadow prices are yield coefficients since adding an exchange reaction $M_i \rightleftharpoons \emptyset$ with a fixed flux $v_\alpha = v_\alpha^{\text{fix}} < 0$ corresponds to adding the metabolite at a rate $|v_\alpha^{\text{fix}}|$.

Let us finally consider the general case $v_\alpha^{\min} \leq v_\alpha \leq v_\alpha^{\max}$. This actually specifies two constraints on the flux, and thus gives rise to two dual variables. One possible interpretation of the resulting dual problem is as follows. In addition to $B_\alpha \leq 0$ one has $B_\alpha \leq \sum_i \pi_i S_{i\alpha}$ rather than a strict equality. A double contribution $B_\alpha v_\alpha^{\min} + (\sum_i \pi_i S_{i\alpha} - B_\alpha) v_\alpha^{\max}$ is added to w . There are also two CS relations, namely $B_\alpha (v_\alpha - v_\alpha^{\min}) = 0$ and $(\sum_i \pi_i S_{i\alpha} - B_\alpha) (v_\alpha^{\max} - v_\alpha) = 0$.

-
- [1] Palsson B Ø 2006 *Systems biology: properties of reconstructed networks* (Cambridge: CUP)
- [2] Temkin O N, Zeigarnik A V and Bonchev D 1996 *Chemical reaction networks: a graph-theoretical approach* (Taylor & Francis, London)
- [3] Sauer U 2004 High-throughput phenomics: experimental methods for mapping fluxomes *Curr. Opin. Biotechnol.* **15** 58
- [4] Schuetz R, Kuepfer L and Sauer U 2007 Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli* *Mol. Systems Biol.* **3** 119
- [5] Almaas E, Kovács B, Vicsek T, Oltval Z N and Barabási A-L 2004 Global organization of the metabolic fluxes in the bacterium *Escherichia coli* *Nature* **427** 839
- [6] Gallos L K, Song C, Havlin S and Makse H A 2007 Scaling theory of transport in complex biological networks *Proc. Natl. Acad. Sci. USA* **104** 7746
- [7] Reed J L, Vo T D, Schilling C H and Palsson B Ø 2003 An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 gsm/gpr) *Genome Biol.* **4** R54.1
- [8] Duarte N C, Herrgård M J and Palsson B Ø 2004 Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model *Genome Res.* doi:10.1101/gr.2250904
- [9] Feist A M, Scholten J C M, Palsson B Ø, Brockman F J and Ideker T 2006 Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri* *Mol. Systems Biol.* doi:10.1038/msb4100046
- [10] Feist A M, Henry C S, Reed J L, Krummenacker M, Joyce A R, Karp P D, Broadbelt J L, Hatzimanikatis V and Palsson B Ø 2007 A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 orfs and thermodynamic information *Mol. Systems Biol.* **3** 121
- [11] Duarte N D, Becker S A, Jamshidi N, Thiele I, Mo M L, Vo T D, Srivas R and Palsson B Ø 2007 Global reconstruction of the human metabolic network based on genomic and bibliomic data *Proc. Natl. Acad. Sci. USA* **104** 1777
- [12] Ibarra R U, Edwards J S and Palsson B Ø 2002 *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth *Nature* **420** 186
- [13] Trawick J D and Schilling C H 2006 Use of constraint-based modeling for the prediction and validation of antimicrobial targets *Biochem. Pharmacol.* **71** 1026
- [14] Zhang C, Shi Z, Gao P, Duan Z and Mao Z 2005 Online prediction of products concentrations in glutamate fermentation using metabolic network model and linear programming *Biochem. Eng. J.* **25** 99
- [15] Burgard A P, Pharkya P and Maranas C D 2003 OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization *Biotech. Bioeng.* **84** 647
- [16] Dantzig G B 1963 *Linear programming and extensions* (Princeton: Princeton University Press)
- [17] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 2007 *Numerical recipes: the art of scientific computing, 3rd edition* (Cambridge: CUP)
- [18] Varma A and Palsson B Ø 1993 Metabolic capabilities of *Escherichia coli*: I. synthesis of biosynthetic precursors and cofactors *J. theor. Biol.* **165** 477
- [19] Varma A and Palsson B Ø 1993 Metabolic capabilities of *Escherichia coli*: II. optimal growth patterns *J. theor. Biol.* **165** 503
- [20] Burgard A P and Maranas C D 2003 Optimization-based framework for inferring and testing hypothesised metabolic objective functions *Biotech. Bioeng.* **82** 670
- [21] Warren P B and Jones J L 2007 Duality, thermodynamics, and the linear programming problem in constraint-based models of metabolism. *Phys. Rev. Lett.* **99** 108101
- [22] Beard D A, Liang S and Qian H 2002 Energy balance for analysis of complex metabolic networks *Biophys. J.* **83** 79
- [23] Beard D A, Babson E, Curtis E and Qian H 2004 Thermodynamic constraints for biochemical networks *J. Theor. Biol.* **228** 327
- [24] Qian H and Beard D A 2005 Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium *Biophys. Chem.* **114** 213
- [25] Smith E B 1993 *Basic chemical thermodynamics* (Oxford: OUP)
- [26] Becker S A, Feist A M, Mo M L, Hannum G, Palsson B Ø and Herrgård M J 2007 Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox *Nature Protocols* **2** 727
- [27] Reed J L and Palsson B Ø 2007 Genome-scale *in silico* models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that compromise network states *Genome* **14** 1797
- [28] Kümmler A, Panke S and Heinemann M 2006 Systematic assignment of thermodynamic constraints in metabolic network models *BMC Bioinformatics* **57** 512
- [29] Mahadevan R and Schilling C H 2003 The effects of alternate optimal solutions in constraint-based genome-scale metabolic models *Metab. Eng.* **5** 264
- [30] Jerrold J H 1999 *Biostatistical analysis* (Upper Saddle River: Prentice Hall)
- [31] Hill B M 1975 A simple general approach to inference about the tail of a distribution *Ann. Stat.* **3** 1163
- [32] Meerschaert M M and Scheffler H P 1998 A simple robust estimation method for the thickness of heavy tails *J. Stat. Plan. Inf.* **71** 19
- [33] Ma H-W and Zeng A-P 2003 The connectivity structure, giant strong component and centrality of metabolic networks *Bioinformatics* **19** 1423
- [34] Tanaka R, Csete M and Doyle J 2005 Highly optimised global organisation of metabolic networks *IEEE Proc. Sys. Biol.* **152** 179
- [35] Zhao J, Yu H, Luo J H, Cao Z W and Li Y X 2006 Hierarchical modularity of nested bow-ties in metabolic networks *BMC Bioinformatics* **7** 386